

Orbit-Product Analysis of (Generalized) Gaussian Belief Propagation

Jason Johnson, Post-Doctoral Fellow, LANL
Joint work with Michael Chertkov and Vladimir Chernyak

Physics of Algorithms Workshop
Santa Fe, New Mexico
September 3, 2009

Overview

Introduction

- ▶ graphical models + belief propagation
- ▶ specialization to Gaussian model

Analysis of Gaussian BP

- ▶ walk-sum analysis for means, variances, covariances¹
- ▶ orbit-product analysis/corrections for determinant²

Current Work on Generalized Belief Propagation (GBP) [Yedidia et al]

- ▶ uses larger “regions” to capture more walks/orbits of the graph (better approximation)
- ▶ *However*, it can also lead to *over-counting* of walks/orbits (bad approximation/unstable algorithm)!

¹Earlier joint work with Malioutov & Willsky (NIPS, JMLR '06).

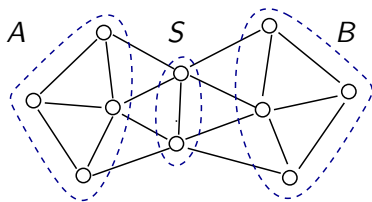
²Johnson, Chernyak & Chertkov (ICML '09).

Graphical Models

A *graphical model* is a multivariate probability distribution that is expressed in terms of interactions among subsets of variables (e.g. pairwise interactions on the edges of a graph G).

$$P(x) = \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{\{i,j\} \in E} \psi_{ij}(x_i, x_j)$$

Markov property:



$$P(x_A, x_B | x_S) = P(x_A | x_S) P(x_B | x_S)$$

Given the potential functions ψ , the goal of *inference* is to compute marginals $P(x_i) = \sum_{x_{V \setminus i}} P(x)$ or the normalization constant Z , which is generally difficult in large, complex graphical models.

Gaussian Graphical Model

Information form of Gaussian density.

$$P(x) \propto \exp \left\{ -\frac{1}{2} x^T J x + h^T x \right\}$$

Gaussian graphical model: sparse J matrix

$$J_{ij} \neq 0 \text{ if and only if } \{i, j\} \in G$$

Potentials:

$$\begin{aligned}\psi_i(x_i) &= e^{-\frac{1}{2} J_{ii} x_i^2 + h_i x_i} \\ \psi_{ij}(x_i, x_j) &= e^{-J_{ij} x_i x_j}\end{aligned}$$

Inference corresponds to calculation of mean vector $\mu = J^{-1} h$, covariance matrix $K = J^{-1}$ or determinant $Z = \det J^{-1}$. Marginals $P(x_i)$ specified by means μ_i and variances K_{ii} .

Belief Propagation

Belief Propagation iteratively updates a set of *messages* $\mu_{i \rightarrow j}(x_j)$ defined on directed edges of the graph G using the rule:

$$\mu_{i \rightarrow j}(x_j) \propto \sum_{x_i} \psi_i(x_i) \prod_{k \in N(i) \setminus j} \mu_{k \rightarrow i}(x_i) \psi(x_i, x_j)$$

Iterate message updates until converges to a fixed point.

Marginal Estimates: combine messages at a node

$$P(x_i) = \frac{1}{Z_i} \underbrace{\psi_i(x_i) \prod_{k \in N(i)} \mu_{k \rightarrow i}(x_i)}_{\tilde{\psi}_i(x_i)}$$

Belief Propagation II

Pairwise Estimates (on edges of graph):

$$P(x_i, x_j) = \frac{1}{Z_{ij}} \tilde{\psi}_i(x_i) \tilde{\psi}_j(x_j) \underbrace{\frac{\psi(x_i, x_j)}{\mu_{i \rightarrow j}(x_j) \mu_{j \rightarrow i}(x_i)}}_{\tilde{\psi}_{ij}(x_i, x_j)}$$

Estimate of Normalization Constant:

$$Z^{\text{bp}} = \prod_{i \in V} Z_i \prod_{\{i, j\} \in G} \frac{Z_{ij}}{Z_i Z_j}$$

BP fixed point is *saddle point* of RHS with respect to messages/reparameterizations.

In trees, BP converges in finite number of steps and is exact (equivalent to variable elimination).

Gaussian Belief Propagation (GaBP)

Messages $\mu_{i \rightarrow j}(x_j) \propto \exp\{\frac{1}{2}\alpha_{i \rightarrow j}x_j^2 + \beta_{i \rightarrow j}x_j\}$.

BP fixed-point equations reduce to:

$$\begin{aligned}\alpha_{i \rightarrow j} &= J_{ij}^2(J_{ii} - \alpha_{i \setminus j})^{-1} \\ \beta_{i \rightarrow j} &= -J_{ij}(J_{ii} - \alpha_{i \setminus j})^{-1}(h_i + \beta_{i \setminus j})\end{aligned}$$

where $\alpha_{i \setminus j} = \sum_{k \in N(i) \setminus j} \alpha_{k \rightarrow i}$ and $\beta_{i \setminus j} = \sum_{k \in N(i) \setminus j} \beta_{k \rightarrow i}$.

Marginals specified by:

$$\begin{aligned}K_i^{\text{bp}} &= (J_{ii} - \sum_{k \in N(i)} \alpha_{k \rightarrow i})^{-1} \\ \mu_i^{\text{bp}} &= K_i^{\text{bp}}(h_i + \sum_{k \in N(i)} \beta_{k \rightarrow i})\end{aligned}$$

Gaussian BP Determinant Estimate

Estimates of pairwise covariance on edges:

$$K_{(ij)}^{\text{bp}} = \begin{pmatrix} J_{ii} - \alpha_{i \setminus j} & J_{ij} \\ J_{ij} & J_{jj} - \alpha_{j \setminus i} \end{pmatrix}^{-1}$$

Estimate of $Z \triangleq \det K = \det J^{-1}$:

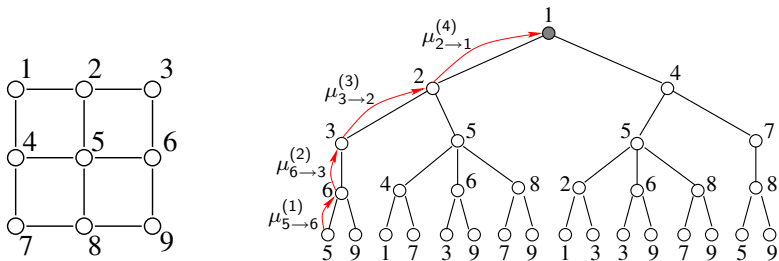
$$Z^{\text{bp}} = \prod_{i \in V} Z_i \prod_{\{i,j\} \in G} \frac{Z_{ij}}{Z_i Z_j}$$

where $Z_i = K_i^{\text{bp}}$ and $Z_{ij} = \det K_{(ij)}^{\text{bp}}$.

Exact in tree models (equivalent to Gaussian elimination),
approximate in loopy models.

The BP Computation Tree

BP marginal estimates are equivalent to the exact marginal in a tree-structured model [Weiss & Freeman].



The BP messages correspond to upwards variable elimination steps in this computation tree.

Walk-Summable Gaussian Models

Let $J = I - R$. If $\rho(R) < 1$ then $(I - R)^{-1} = \sum_{L=0}^{\infty} R^L$.

Walk-Sum interpretation of inference:

$$K_{ij} = \sum_{L=0}^{\infty} \sum_{w:i \xrightarrow{L} j} R^w \stackrel{?}{=} \sum_{w:i \rightarrow j} R^w$$

$$\mu_i = \sum_j h_j \sum_{L=0}^{\infty} \sum_{w:j \xrightarrow{L} i} R^w \stackrel{?}{=} \sum_{w:* \rightarrow i} h_* R^w$$

Walk-Summable if $\sum_{w:i \rightarrow j} |R^w|$ converges for all i, j . Absolute convergence implies convergence of walk-sums (to same value) for arbitrary orderings and partitions of the set of walks. Equivalent to $\rho(|R|) < 1$.

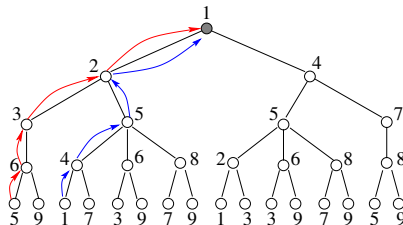
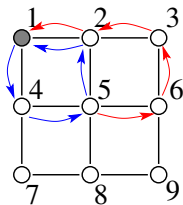
Walk-Sum Interpretation of GaBP

Combine interpretation of BP as exact inference on computation tree with walk-sum interpretation of Gaussian inference in trees:

- ▶ messages represent walk-sums in subtrees of computation tree
- ▶ Gauss BP converges in walk-summable models
- ▶ complete walk-sum for the means
- ▶ incomplete walk-sum for the variances

Complete Walk-Sum for Means

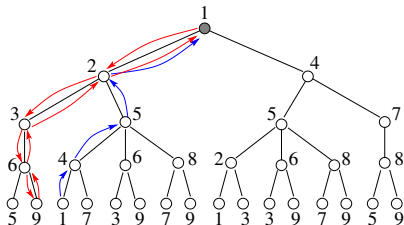
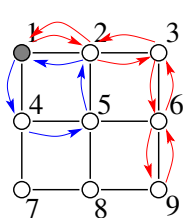
Every walk in \mathcal{G} ending at a node i maps to a walk of the computation tree T_i (ending at root node of T_i)...



Gaussian BP converges to the correct means in WS models.

Incomplete Walk-Sum for Variances

Only those *totally backtracking* walks of \mathcal{G} can be embedded as *closed walks* in the computation tree...



Gaussian BP converges to incorrect variance estimates (underestimate in non-negative model).

Zeta Function and Orbit-Product

What about the determinant?

Definition of Orbits:

- ▶ A walk is *closed* if it begins and ends at same vertex.
- ▶ It is *primitive* if does not repeat a shorter walk.
- ▶ Two primitive walks are *equivalent* if one is a cyclic shift of the other.
- ▶ Define *orbits* $\ell \in \mathcal{L}$ of G to be equivalence classes of closed, primitive walks.

Theorem. Let $Z \triangleq \det(I - R)^{-1}$. If $\rho(|R|) < 1$ then

$$Z = \prod_{\ell} (1 - R^{\ell})^{-1} \triangleq \prod_{\ell} Z_{\ell}.$$

A kind of *zeta function* in graph theory.

Z_{bp} as Totally-Backtracking Orbit-Product

Definition of Totally-Backtracking Orbits:

- ▶ Orbit is *reducible* if it contains backtracking steps $\dots(ij)(ji)\dots$, else it is *irreducible* (or *backtrackless*).
- ▶ Every orbit ℓ has a unique irreducible core $\gamma = \Gamma(\ell)$ obtained by iteratively deleting pairs of backtracking steps until no more remain. Let \mathcal{L}_γ denote the set of all orbits that reduce to γ .
- ▶ Orbit is *totally backtracking* (or *trivial*) if it reduces to the empty orbit $\Gamma(\ell) = \emptyset$, else it is *non-trivial*.

Theorem. If $\rho(|R|) < 1$ then Z^{bp} (defined earlier) is equal to the totally-backtracking orbit-product:

$$Z^{\text{bp}} = \prod_{\ell \in \mathcal{L}_\emptyset} Z_\ell$$

Orbit-Product Correction and Error Bound

Orbit-product correction to Z^{bp} :

$$Z = Z^{\text{bp}} \prod_{\ell \notin \mathcal{L}_\emptyset} Z_\ell$$

Error Bound: missing orbits must all involve cycles of the graph...

$$\frac{1}{n} \left| \log \frac{Z}{Z^{\text{bp}}} \right| \leq \frac{\rho^g}{g(1 - \rho)}$$

where $\rho \triangleq \rho(|R|) < 1$ and g is girth of the graph (length of shortest cycle).

Reduction to Backtrackless Orbit-Product Correction

We may reduce the orbit-product correction to one over just backtrackless orbits γ

$$Z = Z_{\text{bp}} \prod_{\ell} Z_{\ell} = Z_{\text{bp}} \prod_{\gamma} \underbrace{\left(\prod_{\ell \in \mathcal{L}(\gamma)} Z_{\ell} \right)}_{Z'_{\gamma}}$$

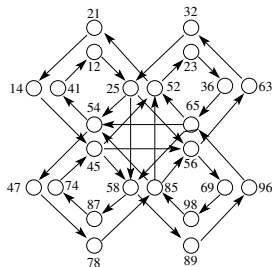
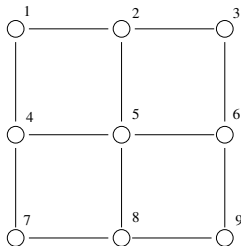
with modified orbit-factors Z'_{γ} based on GaBP

$$Z'_{\gamma} = \left(1 - \prod_{(ij) \in \gamma} r'_{ij} \right)^{-1} \quad \text{where} \quad r'_{ij} \triangleq (1 - \alpha_{i \setminus j})^{-1} r_{ij}$$

The factor $(1 - \alpha_{i \setminus j})^{-1}$ serves to reconstruct totally-backtracking walks at each point i along the backtrackless orbit γ .

Backtrackless Determinant Correction

Define backtrackless graph G' of G as follows: nodes of G' correspond to directed edges of G , edges $(ij) \rightarrow (jk)$ for $k \neq i$.



Let R' be adjacency matrix of G' with modified edge-weights r' based on GaBP. Then,

$$Z = Z_{\text{bp}} \det(I - R')^{-1}$$

Region-Based Estimates/Corrections

Select a set of *regions* $\mathcal{R} \subset 2^V$ that is closed under intersections and cover all vertices and edges of \mathcal{G} .

Define *regions counts* $(n_A \in \mathbb{Z}, A \in \mathcal{R})$ by *inclusion-exclusion rule*:

$$n_A = 1 - \sum_{B \in \mathcal{R} | A \subsetneq B} n_B$$

To capture all orbits covered by any region (without over-counting) we calculate the estimate:

$$Z_{\mathcal{R}} \triangleq \prod_B Z_B^{n_B} \triangleq \prod_B (\det(I - R_B)^{-1})^{n_B}$$

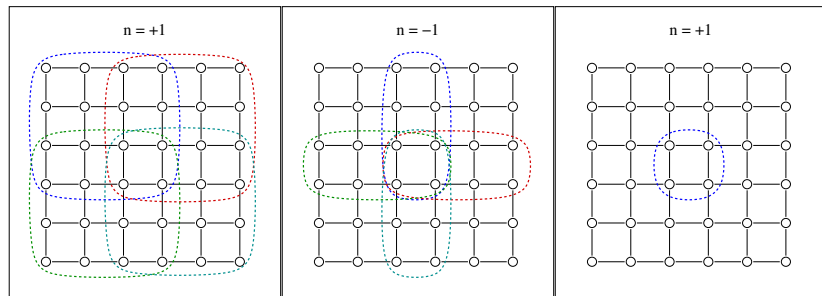
Error Bounds. Select regions to cover all orbits up to length L . Then,

$$\frac{1}{n} \left| \log \frac{Z_B}{Z} \right| \leq \frac{\rho^L}{L(1 - \rho)}$$

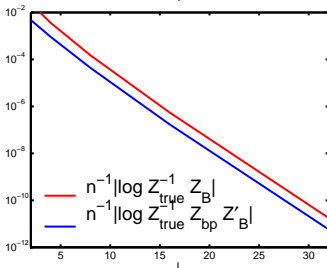
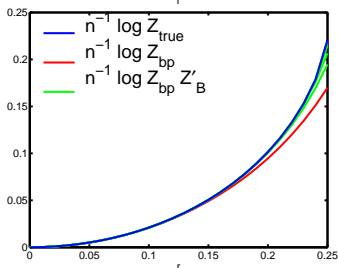
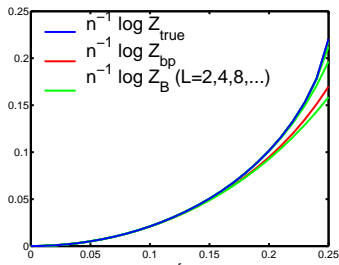
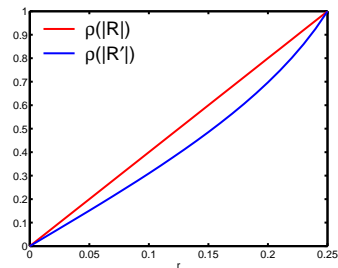
Example: 2-D Grids

Choice of regions for grids: overlapping $L \times L$, $\frac{L}{2} \times L$, $L \times \frac{L}{2}$, $\frac{L}{2} \times \frac{L}{2}$ (shifted by $\frac{L}{2}$).

For example, in 6×6 grid with block size $L = 4$:



256×256 Periodic Grid, uniform edge weights $r \in [0, .25]$.
 Test with $L = 2, 4, 8, 16, 32$.



Generalized Belief Propagation

Select a set of *regions* $\mathcal{R} \subset 2^V$ that is closed under intersections and cover all vertices and edges of \mathcal{G} .

Define *regions counts* ($n_A \in \mathbb{Z}, A \in \mathcal{R}$) by *inclusion-exclusion rule*:

$$n_A = 1 - \sum_{B \in \mathcal{R} | A \subsetneq B} n_B$$

Then, GBP solves for *saddle point* of

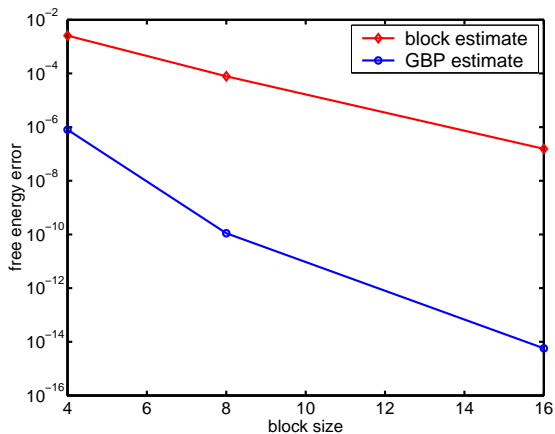
$$Z_{\mathcal{R}}(\psi) \triangleq \prod_{A \in \mathcal{R}} Z(\psi_A)^{n_A}$$

over reparameterizations $\{\psi_A, A \in \mathcal{R}\}$ of the form

$$P(x) = \frac{1}{Z} \prod_{A \in \mathcal{R}} \psi_A(x_A)^{n_A}$$

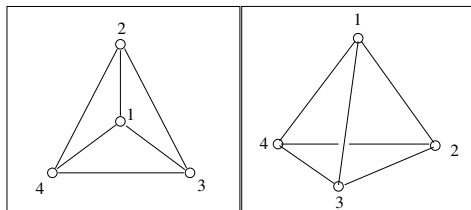
Denote saddle-point by $Z_{\text{gbp}} = Z_{\mathcal{R}}(\psi^{\text{gbp}})$.

Example: 2-D Grid Revisited

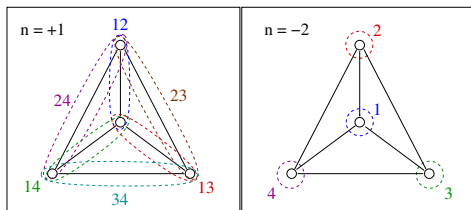


GBP Toy Example

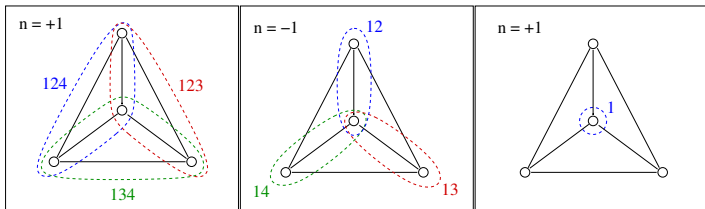
Look at graph $\mathcal{G} = K_4$ and consider different choices of regions...



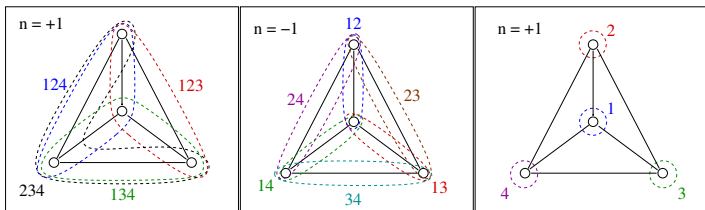
BP Regions:



GBP “3 Δ ” Regions:



GBP “4 Δ ” Regions:



Computational Experiment with equal edge weights $r = .32$ (the model becomes singular/indefinite for $r \geq \frac{1}{3}$).

$$Z = 10.9$$

$$Z_{\text{bp}} = 2.5$$

$$Z_{\text{gbp}}(3\Delta) = 9.9$$

$$Z_{\text{gbp}}(4\Delta) = 54.4!!!$$

GBP with 3Δ regions is big improvement of BP (GBP captures more orbits).

What went wrong with the 4Δ method?

Orbit-Product Interpretation of GBP

Answer: sometimes GBP can *overcount* orbits of the graph.

- ▶ Let $\mathcal{T}(\mathcal{R})$ be the set of hypertrees T one may construct from regions \mathcal{R} .
- ▶ Orbit ℓ *spans* T if we can embed ℓ in T but cannot embed it in any sub-hypertree of T .
- ▶ Let $g_\ell \triangleq \#\{T \in \mathcal{T}(\mathcal{R}) \mid \ell \text{ spans } T\}$.

Orbit-Product Interpretation of GBP:

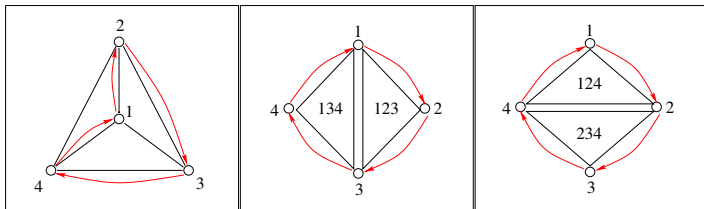
$$Z_{\text{gbp}} = \prod_{\ell} Z_{\ell}^{g_{\ell}}$$

Remark. GBP may also include *multiples of an orbit* as independent orbits (these are not counted by Z).

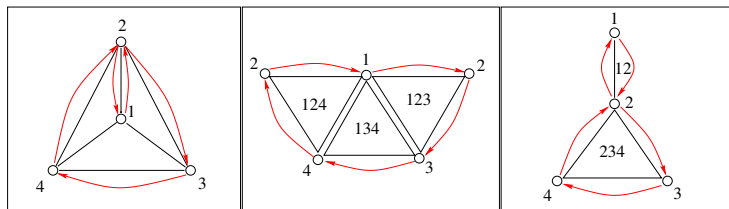
We say GBP is *consistent* if $g_{\ell} \leq 1$ for all (primitive) orbits and $g_{\ell} = 0$ for multiples of orbits (no overcounting).

Examples of Over-Counting

Orbit $\ell = [(12)(23)(34)(41)]:$



Orbit $\ell = [(12)(23)(34)(42)(21)]:$



Conclusion and Future Work

Graphical view of inference in walk-summable Gaussian graphical models that is very intuitive for understanding iterative inference algorithms and approximation methods.

Future Work:

- ▶ many open questions on GBP.
- ▶ multiscale method to approximate longer orbits from coarse-grained model.
- ▶ beyond walk-summable?